

FORM PTO-1390 (Modified)
(REV 11-2000)

U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

ATTORNEY'S DOCKET NUMBER

**TRANSMITTAL LETTER TO THE UNITED STATES
DESIGNATED/ELECTED OFFICE (DO/EO/US)
CONCERNING A FILING UNDER 35 U.S.C. 371**

A32312-PCT USA

U.S. APPLICATION NO. (IF KNOWN, SEE 37 CFR

09/913746

INTERNATIONAL APPLICATION NO.
PCT/US00/04505

INTERNATIONAL FILING DATE
22 February 2000

PRIORITY DATE CLAIMED
19 February 1999

TITLE OF INVENTION

CUT AND PASTE DOCUMENT SUMMARIZATION SYSTEM AND METHOD

APPLICANT(S) FOR DO/EO/US

MCKEOWN, Kathleen R. and JING, Hongyan

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☐ This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (24) indicated below.
4. ☒ The US has been elected by the expiration of 19 months from the priority date (Article 31).
5. ☒ A copy of the International Application as filed (35 U.S.C. 371 (c) (2))
 - a. ☒ is attached hereto (required only if not communicated by the International Bureau).
 - b. ☐ has been communicated by the International Bureau.
 - c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
 - a. ☐ is attached hereto.
 - b. ☐ has been previously submitted under 35 U.S.C. 154(d)(4).
- ☒ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371 (c)(3))
 - a. ☐ are attached hereto (required only if not communicated by the International Bureau).
 - b. ☐ have been communicated by the International Bureau.
 - c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
 - d. ☒ have not been made and will not be made.
- ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
- ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371 (c)(4)).
- ☐ An English language translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371 (c)(5)).
- ☒ A copy of the International Preliminary Examination Report (PCT/IPEA/409).
12. ☒ A copy of the International Search Report (PCT/ISA/210).

Items 13 to 20 below concern document(s) or information included:

13. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
14. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
15. ☐ A **FIRST** preliminary amendment.
16. ☐ A **SECOND** or **SUBSEQUENT** preliminary amendment.
17. ☐ A substitute specification.
18. ☐ A change of power of attorney and/or address letter.
19. ☐ A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.
20. ☐ A second copy of the published international application under 35 U.S.C. 154(d)(4).
21. ☐ A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).
22. ☒ Certificate of Mailing by Express Mail
23. ☒ Other items or information:

Form PCT/IB/301; Form PCT/IB/304; Form PCT/IPEA/401; Amendment Under Article 34; Form PCT/IPEA/402; form PCT/IB/308; Form PCT/IB/332; Form PCT/IPEA/408; Response to Written Opinion; a postcard and a check in the amount of \$538.00.

Express Mail Label No.: EK839853032US

Date of Deposit: August 16, 2001

U.S. APPLICATION NO. (IF KNOWN, SEE 37 CFR 1.53) 09/913746		INTERNATIONAL APPLICATION NO. PCT/US00/04505		ATTORNEY'S DOCKET NUMBER A32312-PCT USA	
---	--	--	--	---	--

24. The following fees are submitted: BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)) : <input type="checkbox"/> Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO \$1000.00 <input type="checkbox"/> International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO \$860.00 <input type="checkbox"/> International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO \$710.00 <input checked="" type="checkbox"/> International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) \$690.00 <input type="checkbox"/> International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) \$100.00 ENTER APPROPRIATE BASIC FEE AMOUNT =				CALCULATIONS PTO USE ONLY	
				\$690.00	
Surcharge of \$130.00 for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492 (e)).				\$0.00	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total claims	37 - 20 =	17	x \$18.00	\$306.00	
Independent claims	4 - 3 =	1	x \$80.00	\$80.00	
Multiple Dependent Claims (check if applicable). <input type="checkbox"/>				\$0.00	
TOTAL OF ABOVE CALCULATIONS =				\$1,076.00	
<input checked="" type="checkbox"/> Applicant claims small entity status. (See 37 CFR 1.27). The fees indicated above are reduced by 1/2.				\$538.00	
SUBTOTAL =				\$538.00	
Processing fee of \$130.00 for furnishing the English translation later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492 (f)).				\$0.00	
TOTAL NATIONAL FEE =				\$538.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31) (check if applicable). <input type="checkbox"/>				\$0.00	
TOTAL FEES ENCLOSED =				\$538.00	
				Amount to be:	\$
				refunded	
				charged	\$

a. ☒ A check in the amount of **\$538.00** to cover the above fees is enclosed.


b. ☐ Please charge my Deposit Account No. _____ in the amount of _____ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. **02-4377**. A duplicate copy of this sheet is enclosed.

d. ☐ Fees are to be charged to a credit card. **WARNING:** Information on this form may become public. **Credit card information should not be included on this form.** Provide credit card information and authorization on PTO-2038.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

Henry Tang BAKER BOTTS LLP 30 Rockefeller Plaza New York, NY 10112-0228	<div style="text-align: center;">  SIGNATURE </div> <div style="text-align: center;"> Paul D. Ackerman NAME </div> <div style="text-align: center;"> 39,891 REGISTRATION NUMBER </div> <div style="text-align: center;"> August 16, 2001 DATE </div>
--	---

CUT AND PASTE DOCUMENT SUMMARIZATION
SYSTEM AND METHOD

Statement of Government Rights

5 The United States Government may have certain rights to the invention set forth herein pursuant to a grant by the National Science Foundation, Contract No. IRI-96-198124

Statement of Related Applications

10 This application claims the benefit of United States provisional patent application, Serial No. 60/120,657, entitled "Summary Generation Through Intelligent Cutting and Pasting of the Input Document" which was filed on February 19, 1999.

Field of the Invention

The present invention relates generally to information summarization and more particularly relates to systems and methods for generating a summary of a document using automated cutting and pasting of the input document.

15 **Background of the Invention**

The amount of information available today drastically exceeds that of any other time in history. With the continuing expansion of the Internet, this trend will likely continue well into the future. Often, people conducting research of a topic are faced with information overload as the number of potentially relevant documents exceeds the researcher's ability to individually review each document. To address this problem, information summaries are often relied on by researchers to quickly evaluate a document to determine if it is truly relevant to the problem at hand.

25 Given the vast collection of documents available, there is interest in developing and improving the systems and methods used to summarize information content. For individual documents, domain-dependent template based systems and

domain-independent sentence extraction methods are known. Such known systems can provide a reasonable summary of a single document when the domain is known.

Many presently available summarizers extract sentences from the original documents to produce summaries. However, since the sentences are
5 generally extracted without supporting context information, the resulting summaries can be incoherent, and in some cases, can convey misleading information.

Therefore, there remains a need for systems and methods which can generate a more readable and concise summary of a document.

Summary of the Invention

10 It is an object of the present invention to provide a system and method for generating a summary of a document.

It is another object of the present invention to provide a summarization system which extracts sentences from an input document and then transforms the extracted sentences such that a concise, coherent and accurate summary results.

15 It is a further object of the present invention to provide a system and method for generating a summary of a set document which use automated cutting and pasting of the input document.

A present method for generating a summary of an input document includes extracting at least one sentence from the document. The extracted sentences
20 are parsed into components, preferably in a parse tree representation. Sentence reduction is performed to mark components which can be removed from the extracted sentences. Sentence combination is performed to mark components of two or more sentences which can be merged. Sentence combination also includes a paste operation to operate on the marked components to effect the indicated removal and combination
25 of sentence components.

A preferred sentence reduction operation includes measuring the contextual importance of the components; measuring the probabilistic importance of the components based on a given corpus; measuring the importance of the components based on linguistic knowledge; synthesizing the contextual, probabilistic
30 and knowledge based importance measures into a relative importance score for each

20240924 09:46:04

component; and marking for removal those components with an importance score below a threshold value.

5 The contextual importance can be measured by establishing a plurality of lexical links of at least one type among the components in a local context in the document and computing a context importance score according to the type, number and direction of lexical links associated with each component. The types of lexical links can include repetition, inflectional variants, derivational variants, synonyms, hypernyms, antonyms, part-of, entailment, and causative links.

10 In a preferred method, the sentence combination operation includes identifying sentence combination operations from a sentence combination subcorpus and developing rules regarding the application of the sentence combination operations. The combination rules are then applied to the extracted sentences after sentence reduction to identify and merge suitable sentences from the original article. The sentence combination operations can be selected from the group including add
15 descriptions, aggregations, substitute incoherent phrases, substitute phrases with more general or more specific information, and mixed operations.

A present system for generating a summary of an input document includes an extraction module which receives the input document and extracts at least one sentence related to a focus of the document. A summary sentence generation
20 module is provided, which generally includes a sentence reduction module and a sentence combination module. The system includes a grammatical parser operatively coupled to the generation module for parsing the extracted sentences into components in a grammatical representation. A combined lexicon and a corpus of human generated summaries are operatively coupled to the generation module for use by the
25 operational modules during summary generation.

The corpus can further include a sentence generation subcorpus and a sentence reduction subcorpus. The subcorpora can be generated manually or through the use of a decomposition module.

30 Preferably, the sentence reduction module is cooperatively engaged with the combined lexicon and performs context importance processing on the components of the grammatical representation. Context importance processing can

include establishing a plurality of lexical links of at least one type for the components and generating a context importance score based on the type and number of links associated with the components. The number and type of lexical links can vary, however a preferred set of lexical link types includes repetition, inflectional variants, derivational variants, synonyms, hypernyms, antonyms, part-of, entailment, and causative links.

Preferably, the sentence reduction module further computes the relative importance of the components based on linguistic knowledge stored in the combined lexicon. The sentence reduction module can also be cooperatively engaged with the corpus and perform probabilistic importance processing on the components of the grammatical representation in accordance with the particular corpus used.

The sentence combination module can be used to identify sentence combination operations from a sentence combination subcorpus and develop rules regarding the application of the sentence combination operations. The combination module applies the combination rules to the extracted sentences after sentence reduction to identify and merge suitable sentences from the original article.

A decomposition module in accordance with the present application can be used to evaluate human generated summaries and map corresponding portions of the summaries to the original documents. The decomposition module indexes words in the summary and the original document. A Hidden Markov Model is then built based on heuristic rules to determine the probability of phrases in the summary sentence matching a given phrase in the original document. A Viterbi algorithm can then be employed to determine the best solution for the Hidden Markov Model and generate a mapping between summary phrases and the original document. This mapping can be used to generate, among other things, a sentence reduction subcorpus and a sentence combination subcorpus. Such a decomposition module can be operatively coupled to the corpus in the summary generation system described above.

Brief Description of the Drawing

Further objects, features and advantages of the invention will become apparent from the following detailed description taken in conjunction with the accompanying figures showing illustrative embodiments of the invention, in which

5 Figure 1 is a block diagram of the system architecture of the present document summarization system;

 Figure 2 is a flow chart illustrating an exemplary embodiment of a sentence reduction operation in accordance with the summarization system of Figure 1;

10 Figure 3 is a pictorial diagram of an exemplary parse tree sentence representation;

 Figure 4 is a flow chart illustrating an exemplary embodiment of a sentence combination operation in accordance with the present summarization system of Figure 1;

15 Figure 5 is a table illustrating exemplary sentence combination operations for the sentence combination operation of Figure 4;

 Figure 6 is a table illustrating exemplary sentence combination rules for applying the sentence combination operations of Figure 5;

20 Figure 7 is a flow diagram illustrating the operation of the corpus decomposition module of Figure 1; and

 Figure 8 is a pictorial diagram of a Hidden Markov Model for use in a corpus decomposition module.

25 Throughout the figures, the same reference numerals and characters, unless otherwise stated, are used to denote like features, elements, components or portions of the illustrated embodiments. Moreover, while the subject invention will now be described in detail with reference to the figures, it is done so in connection with the illustrative embodiments. It is intended that changes and modifications can be made to the described embodiments without departing from the true scope and
30 spirit of the subject invention as defined by the appended claims.

Detailed Description of Preferred Embodiments

The present summarization systems and methods generate a generic, domain-independent single-document summary of a received input document. Figure 1 is a block diagram illustrating the system architecture of an exemplary embodiment of the present summarization system. Such a system can be implemented on various computer hardware, software and operating system platforms. The particular system components selected are not critical to the practice of the present invention. For example, the present system of Figure 1, can be implemented on a personal computer system, such as an IBM compatible system.

Referring to Figure 1, an input document 105 in computer readable form is applied to an extraction module 110 which determines the focus of the document 105 and extracts sentences from the document accordingly. A number of extraction techniques can be used in the extraction module 110. In a preferred embodiment, the extraction module 110 links words in a sentence to other words in the input document 105 through repetitions, morphological relations and lexical relations. An importance score can then be computed for each word in the article 105 based on the number, type and direction (forward, backward) of the lexical links associated with the word. A sentence score can be determined by adding the importance score for each of the words in the sentence and normalizing the sum based on the number of words in the sentence. The sentences can then be extracted based on the highest relative sentence scores.

The extraction module 110 provides the extracted sentences 115 to a generation module 120. The generation module 120 also receives the original document 105 as an input. The generation module 120 further includes a sentence reduction module 135 and a sentence combination module 140. The sentence reduction module 135 provides a marked up parse tree as input data for the sentence combination module 140, which generates and outputs the summary sentences.

The generation module 120 is operatively coupled to a corpus of human-written summaries 165, a lexical database 170, and a combined reusable lexicon 175.

The corpus 165 generally includes a broad collection of human-generated summaries as well as the corresponding original documents. The corpus 165 can also include a sentence reduction subcorpus 165a and a sentence combination subcorpus 165b which can be generated manually or through a decomposition module. The sentence reduction subcorpus 165a includes entries of sentence pairs linking an original sentence to a human reduced sentence. The sentence combination subcorpus 165b includes mappings from human combined sentences to two or more original sentences.

A suitable exemplary corpus 165 was generated using *Communications-related Headlines*, a free daily online news service provided by the Benton Foundation (<http://www.benton.org>). The articles from this particular service are communication related, but the topics involved are very broad, including law, company mergers, new technologies, labor issues and so on. Of course, other sources of document summaries can also be used to generate a suitable corpus. To insure that the resulting corpus is somewhat generic, the articles from the selected source should not possess a particular writing style. Thus, preferred sources feature articles from multiple sources or articles from various sections of one or more source. A suitable corpus 165 was generated in four major steps. First, human-written, single document summaries are received from the source. Second, the original documents are retrieved and correlated to the respective summary. The retrieved documents are then "cleaned" by removing irrelevant material such as indexes and advertisements. Finally, the quality of the correspondence between the summary and the original document is verified. The cleaning and verification processes are generally performed manually. The sentence reduction subcorpus 165a and sentence combination subcorpus 165b entries were generated by the decomposition module 185, the operation of which is explained below.

The lexical database 170 can take the form of the WordNet database, which is described in the article "WordNet: A lexical Database for English", by G.A. Miller, Communications of the ACM, Vol. 38, No. 11. pp. 39-41, November 1995. A suitable embodiment of the combined lexicon 175 can be constructed by combining multiple, large-scale resources such as WordNet, the English Verb Classes and

Alternations (EVCA) database, the COMLEX syntax dictionary and the Brown Corpus tagged with WordNet senses. The combined lexicon 175 can be formed by encoding the EVCA database with COMLEX compatible syntax and merging the EVCA into the COMLEX database. This results in each verb in the combined
5 lexicon 175 being marked with a list of subcategorizations and alternate syntactic patterns. Preferably, WordNet is added to the EVCA/COMLEX combination to refine the syntactic information and provide additional lexical information to the lexicon 175.

The generation module 120 is also cooperatively coupled to natural
10 language processing (NLP) tools such as a syntactic parser 180 and a co-reference resolving module 190 which can include anaphora resolution. These tools can be software modules which are called by the generation module 120. A suitable syntactic parser 180 is the English Slot Grammar (ESG) parser available from International Business Machines, Inc. A suitable co-reference resolving module 190
15 is the Deep Read system, available from Mitre, Inc.

Figure 2 is a flow diagram further illustrating the operation of the sentence reduction module 135. The reduction module 135 receives extracted sentences 115 as input (step 205). The reduction module invokes the parser 180 to grammatically parse the extracted sentences 115 and generate a parse tree
20 representation of the sentences (step 210). In step 215 contextual importance is determined by detecting lexical links among words in a local context and then computing an importance score based on the number, type and direction of lexical links detected. The context processing step 215 generates an importance score for each node in the parse tree indicating the relative importance of the nodes to the focus
25 of the input document 105.

The number, type and direction (forward, backward) of lexical links used in the practice of the present invention may vary. An empirical study has demonstrated that the following nine lexical relation types provide a meaningful representation of contextual importance: (1) repetition, (2) inflectional variants, (3)
30 derivational variants, (4) synonyms, (5) hypernyms, (6) antonyms, (7) part-of, (8) entailment (for example: *kill* → *die*), and (9) causative (for example: *eat* → *chew*).

Inflectional variants (2) and derivational variants can be derived from the CELEX database content, available from the Centre for Lexical Information, Max Planck Institute for Psycholinguistics, Nijmegen, which can be in the combined lexicon 175. The other lexical relations can be extracted using the separate lexical database 170, such as WordNet. To frame the local context of a word, a number of sentences before and after the current sentence location are evaluated for the presence of lexical links. The number of sentences selected for this operation involves balancing the level of contextual depth to the amount of processing overhead. Using the five sentences before and the five sentences after the current sentence has been found to provide reasonable local context without incurring excessive processing overhead.

After the lexical links have been identified (step 215 a), an importance score for each word in the extracted sentences can be calculated (step 215 b). Lexical links from the current sentence to subsequent sentences are referred to as forward links and those from the current sentence to preceding sentences are referred to as backward links. The importance score, referred to as the context weight, can be computed as follows:

- 1) $ForwardWeight(w) = \sum_{i=1}^9 (WixLi(w))$
- 2) $BackwardWeight(w) = \sum_{i=1}^9 (WixBnumi(w))$
- 3) $TotalWeight(w) = ForwardWeight(w) + BackwardWeight(w)$
- 4) $Ratio(w) = \frac{\max(ForwardWeight(w), BackwardWeight(w))}{TotalWeight(w)}$
- 5) $ContextWeight = Ratio(w) \times TotalWeight(w)$

where $ForwardWeight(w)$ computes the weight of forward links, $BackwardWeight(w)$ computes the weight of backward links, $TotalWeight(w)$ represents the sum of all links and $Ratio(w)$ computes a weight for the location of the word. To compute the weight of various lexical links, each type of link is assigned a weighted value according to its relative importance. For example, the nine lexical relations set forth

above were presented in descending order of importance and accordingly can be assigned linearly decreasing weights such as (1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2).

The value of $Ratio(w)$ represents the value assigned based on the location of the word in the original document. For example, when a sentence introduces a topic or ends a topic, it is considered more important and the components of those sentences will be assigned a relatively higher location value.

The use of various types of lexical relations improves the relatedness of a word to the main topic. Although simple relations like repetition and synonymy can be used to determine a measure of contextual importance, these surface relations are generally unable to detect more subtle connections between words.

Following context processing (step 215) the reduction module 135 can perform interdependency processing using a probability analysis based on the corpus 165 of human-written reduction based sentences. Such an analysis can indicate the degree of correlation between components in a sentence, such as the relationship between a verb and its subclause.

The probability computation can be performed based on parse trees using probabilities to indicate the degree of correlation between a parent node and its child nodes in the parse tree. Figure 3 illustrates an exemplary fragment of a parse tree used to explain the operation of the probability computation. In Fig. 3, The main verb "give" is the parent nodes 300, and it has four children nodes: subclause conjunct 305, subject 310, indirect object 315 and object 320, respectively. The parse tree can also include further levels below the children nodes, such as nodes *ndet* 325 and *adjp* 330 below child node *obj* 320 and nodes *lconj* 335 and *rconj* 340 below node *adjp* 330, respectively.

To measure the interdependency between the verb *give* and its subclause 305, the probability that the subclause is removed when the verb is *give*, can be represented by $PROB("when_clause\ is\ removed" | verb = give)$. This conditional probability is transformed using Bayes's rule to:

$$PROB("when_clause\ is\ removed" | v = give) = \frac{PROB("v = give" | "when_clause\ removed") * PROB("when_clause\ is\ removed")}{PROB(verb = give)}$$

In a similar fashion, the probabilities that a clause will be reduced or remain unchanged can be calculated in a similar manner.

The probability associated with the other child nodes from the current root node is calculated in a similar manner. After the probabilities for each of the first level child nodes is calculated, each of the child nodes in the current level of the tree is then treated as a parent node and the process is repeated through each descending level of the parse tree until every parent-child node pair has been considered. The probabilities for the corpus 165 can be calculated and stored in a look-up table which is used when a reduction module 135 is run.

The context processing of step 215 and probability processing of step 220 provide a relative ranking of sentence components. However, this ranking does not necessarily provide a measure of which components be included to provide a grammatically correct summary sentence. Thus, preferably, after the probability analysis of step 220, reduction processing based on linguistic knowledge is performed (step 225). In this operation, the reduction module 135 works in cooperation with the combined lexicon 175.

The linguistic knowledge processing step 225 operates with the combined lexicon 175 to evaluate the parse tree for each extracted sentence 115 and determine which children nodes are essential to maintain the grammatical correctness of the component represented by the parent node. Linguistic judgments are identified in the parse tree by assigning a binary tag to each node in the parse tree. The value of a tag is either *essential* or *reducible*, indicating whether or not a node is indispensable to its parent node. For example, referring to Figure 3, the lexicon 175 will indicate that the verb *give* needs a subject and two objects. Thus the child nodes *subj* 310, *iobj* 315 and *obj* 320 can be marked as essential. In this case, the child node subclause 305 is then rendered non-essential and will be marked as reducible. The lexicon 175 can also include collocations, such as *consist of* or *replace with*, which prevents removal of indispensable components.

Once the linguistic knowledge processing is applied in step 225, a reduction operation (step 230) can take place. The reduction operation process can be viewed as a series of decision making steps along the edges of a parse tree. Beginning

with the root node of the parse tree, the immediate child nodes are evaluated to determine which child nodes can be removed. A child node can be removed if three conditions are satisfied. The first condition is that the component is not a local focus. To determine whether a component is a local focus, the ratio of the context importance score (step 215b) of the child node to that of the root node is calculated. The child node is then considered unimportant if the calculated ratio is smaller than a threshold value. The second condition is that the corpus probability value (step 220) indicating that the special syntactic component of the root is removed is higher than a threshold. The final condition is that the linguistic analysis in step 225 indicates that the child node as reducible.

When the conditions to remove a child node are satisfied, the child node is tagged as "removable" and processing on that branch of the tree terminates. For the child nodes which are retained, the lower levels of the parse tree are evaluated by repeating this process in a similar manner through the tree. The reduction operation step 230 is complete when there are no more nodes to consider. This also concludes processing of the sentence reduction module and results in the parse trees being marked with those components which can be removed or altered by the subsequent paste module 150 operation.

Following processing by the sentence reduction module 135, processing by the sentence combination module 140 is performed. The operation of the sentence combination module 140 is further illustrated in the flow chart of Figure 4.

Using the sentence combination subcorpus 165b, the sentence combination module evaluates the extracted sentence to identify applicable sentence combination operations (step 410). Figure 5 is a table illustrating combination operations such as: add descriptions 510, aggregations 515, substitute incoherent phrases 520, substitute phrases with more general or more specific information 525 and mixed operations 530.

From the sentence combination subcorpus 165b, sentence combination rules are also established to determine whether and how the sentence combination operations of step 410 will take place (step 415). The result is a set of sentence

combination rules 420, such as those set forth in Figure 6. The rules illustrated in Figure 6 are exemplary and non-exhaustive. These sentence combination rules 420 were determined empirically by manual inspection of the sentence combination subcorpus 165b. Using the input article 105 and the extracted sentences reduced by the sentence reduction module 135 the sentence combination module 140 in cooperation with the co-reference resolution module 190 applies the sentence combination rules 420 (step 425). The result of step 425 is that the parse trees of the sentences being combined are appropriately tagged to effect the sentence combination. The combination operation is then realized in step 430 using a tree adjoining grammar (TAG) formalism, as described by A. Joshi, "Introduction to Tree-Adjoining Grammars," in Mathematics of Language, John Benjamins, Amsterdam, 1987. In this way, the sentence combination module 140 performs a paste operation on the marked parse trees and generates a summary sentence.

The document summary is generated by combining the summary sentences. The most straight forward combination is to maintain the order of sentences as they were extracted, however, other sequencing arrangements can also be employed.

As noted above in connection with Figure 1, the corpus decomposition module 185 operates on the corpus 165 to generate the sentence reduction subcorpus 165a and the sentence combination subcorpus 165b. The decomposition module 185 generally operates to evaluate the human written summaries in the corpus 165, compare the summary sentences to the original document, determine if a summary sentence was generated by a cut and past operation and identify where the components of the summary sentences were taken from in the original documents. The operation of the decomposition module 185 is illustrated in the flow diagram of Figure 7.

Referring to Figure 7, the decomposition module 185 uses the human-generated summary and original document as inputs to an indexing operation (step 705). During indexing, each word in the original document is indexed according to its positions in the original document. A convenient way of referencing these occurrences is by sentence number and word number in the original document.

To evaluate the index of words, a set of heuristic rules is developed by manual inspection of the corpus 165. Such inspection reveals that human-generated summaries often include one or more of six operations: sentence reduction, sentence combination, syntactic transformation, lexical paraphrasing, generalization/specification, and content reordering. The heuristic rules can be represented using a bigram probability $PROB(W_2 = (S_2, W_2) | W_1 = (S_1, W_1))$ (abbreviated as $PROB(W_2 | W_1)$ in the following discussion). The probability values can be assigned in the following manner:

● IF $((S_1 = S_2) \text{ and } (W_1 = W_2 - 1))$ (i.e., the words are in two adjacent positions in the document), THEN $PROB(W_2 | W_1)$ is assigned the maximal value, P1. (Rule: Two adjacent words in the summary are most likely to come from two adjacent words in the document.)

● IF $((S_1 = S_2) \text{ and } (W_1 < W_2 - 1))$, THEN $PROB(W_2 | W_1)$ is assigned the second highest value, P2. (Rule: Adjacent words in the summary are highly likely to come from the same sentence in the document, retaining their relative precedent relation, as in sentence reduction. This rule can be further refined by adding restrictions on distance between words.)

● IF $((S_1 = S_2) \text{ and } (W_1 > W_2))$, THEN $PROB(W_2 | W_1)$ is assigned the third highest value, P3. (Rule: Adjacent words in the summary are likely to come from the same sentence in the document but reverse their relative orders, such as in the case of sentence reduction with syntactic transformations.)

● IF $(S_2 - CONST < S_1 < S_2)$, THEN $PROB(W_2 | W_1)$ is assigned the fourth highest value, P4. (Rule: Adjacent words in the summary can come from nearby sentences in the document and retain their relative order, such as in sentence combination. CONST is a small constant such as 3 or 5.)

● IF $(S_2 < S_1 < S_2 + CONST)$, THEN $PROB(W_2 | W_1)$ is assigned the fifth highest value, P5. (Rule: Adjacent words in the summary can come from nearby sentences in the document but reverse their relative orders.)

● IF $(|S_2 - S_1| \geq CONST)$ THEN $PROB(W_2 | W_1)$ is assigned a small value, P6. (Rule: Adjacent words in the summary are not very likely to come from sentences far apart.)

Based on the above heuristic principles, a Hidden Markov Model can be generated, such as is illustrated in Figure 8 (step 710). The nodes in the Hidden Markov Model represent possible positions in the document, and the edges output the probability of going from one node to another. This Hidden Markov Model is used in finding the most likely position sequence in a subsequent processing operation.

Assigning values to P1-P6 is performed empirically. For example, the maximal value can be assigned 1 and others are assigned evenly decreasing values 0.9, 0.8 and so on. The order of the above rules is based on the empirical observations on a particular set of summaries. These values, however, can be adjusted or even trained for different corpora.

A Viterbi algorithm can be used to evaluate the Hidden Markov Model and find the most likely sequence of words incrementally (step 715). The Viterbi algorithm first finds the most likely sequence for (*Word*₁*Word*₂), for each possible position of *Word*₂. This information is then used to compute the most likely sequence for (*Word*₁*Word*₂*Word*₃), for each possible position of *Word*₃. The process repeats until all the words in the sequence have been considered.

After evaluation by the Viterbi algorithm, post-editing operations can be used to cancel mismatches that occur in the corpus analysis. The result is that summary sentences are matched to the corresponding phrases in the document. Once the summary sentences are so matched, it is a simple endeavor to sort the various matchings to one of the sentence reduction subcorpus 165a and sentence combination subcorpus 165b. In addition, the decomposition module 185 can be used as a stand alone tool, apart from the rest of the present summary generation system, to perform various summary analysis operations.

Although the present invention has been described in connection with specific exemplary embodiments, it should be understood that various changes, substitutions and alterations can be made to the disclosed embodiments without departing from the spirit and scope of the invention as set forth in the appended claims.

CLAIMS

1. A system for generating a summary of an input document comprising:
an extraction module, the extraction module receiving the input
document and extracting at least one sentence related to a focus of the document;
5 a summary sentence generation module operatively coupled to the
extraction module;
a grammatical parser operatively coupled to the generation module for
parsing the extracted sentences into components in a grammatical representation;
a combined lexicon operatively coupled to the generation module; and
10 a corpus of human generated summaries operatively coupled to the
generation module.
2. The system for generating a summary of an input document of claim 1,
wherein the generation module further comprises a sentence reduction module.
3. The system for generating a summary of an input document of claim 2,
15 wherein the sentence reduction module is cooperatively engaged with the corpus and
performs probabilistic importance processing on the components of the grammatical
representation in accordance with the corpus.
4. The system for generating a summary of an input document of claim 3,
wherein the sentence reduction module is cooperatively engaged with the combined
20 lexicon and performs context importance processing on the components of the
grammatical representation.
5. The system for generating a summary of an input document of claim 4,
wherein the context importance processing includes establishing a plurality of lexical
links of a least one type for the components and generating a context importance score
25 based on the type and number of links associated with the components.

6. The system for generating a summary of an input document of claim 5, wherein the sentence reduction module further computes the relative importance of the components based on linguistic knowledge stored in the combined lexicon.
7. The system for generating a summary of an input document of claim 1,
5 wherein the generation module further comprises a sentence combination module.
8. The system for generating a summary of claim 7, wherein the sentence combination module is operatively coupled to the corpus and wherein the sentence combination module:
- 10 identifies at least one sentence combination operation;
establishes at least one rule for applying the sentence combination operation; and
applies the at least one rule to combine at least two extracted sentences.
9. The system for generating a summary of claim 8, wherein the at least one
15 sentence combination operation is selected from the group consisting of add descriptions, aggregations, substitute incoherent phrases, substitute phrases with more general or more specific information, and mixed operations.
10. The system for generating a summary of claim 9, wherein the at least one rule
20 to combine extracted sentences includes replacing a partial name phrase with a full name phrase.
11. The method of generating a summary of claim 10, wherein the at least one rule
25 to combine extracted sentences includes determining if two sentences having a common subject are proximate and whether at least one sentence is marked for reduction then removing the subject of the second sentence and combining with the first sentence using the connective "and."

12. The system for generating a summary of an input document of claim 1, wherein the generation module further comprises a sentence reduction module and a sentence combination module.

13. The system for generating a summary of an input document of claim 12, wherein the sentence reduction module is cooperatively engaged with the combined lexicon and performs context importance processing on the components of the grammatical representation.

14. The system for generating a summary of an input document of claim 13, wherein the context importance processing includes establishing a plurality of lexical links of a least one type for the components and generating a context importance score based on the type and number of links associated with the components.

15. The system for generating a summary of an input document of claim 14, wherein the sentence reduction module further computes the relative importance of the components based on linguistic knowledge stored in the combined lexicon.

16. The system for generating a summary of an input document of claim 15, wherein the sentence reduction module is cooperatively engaged with the corpus and performs probabilistic importance processing on the components of the grammatical representation in accordance with the corpus.

17. The system for generating a summary of an input document of claim 12, wherein the sentence combination module is operatively coupled to the corpus and wherein the sentence combination module:

identifies at least one sentence combination operation;

establishes at least one rule for applying the sentence combination operation; and

applies the at least one rule to combine at least two extracted sentences.

18. The system for generating a summary of claim 17, wherein the at least one sentence combination operation is selected from the group consisting of add descriptions, aggregations, substitute incoherent phrases, substitute phrases with more general or more specific information, and mixed operations.

5 19. The system for generating a summary of claim 18, wherein the at least one rule to combine extracted sentences includes replacing a partial name phrase with a full name phrase.

10 20. The method of generating a summary of claim 19, wherein the at least one rule to combine extracted sentences includes determining if two sentences having a common subject are proximate and whether at least one sentence is marked for reduction then removing the subject of the second sentence and combining with the first sentence using the connective "and."

15 21. The system for generating a summary of an input document of claim 1, further comprising a decomposition module operatively coupled to the corpus, the decomposition module analyzing the corpus and generating a sentence reduction subcorpus and a sentence combination subcorpus.

20 22. A method of generating a summary of an input document comprising:
extracting at least one sentence from the document;
parsing the at least one sentence into components;
20 performing a sentence reduction operation to mark components which can be removed from the sentence;
performing a sentence combination operation to mark components of at least two sentences which can be merged; and
operating on the marked components to effect the indicated removal
25 and combination of sentence components.

23. The method of generating a summary of claim 22, wherein the sentence reduction operation comprises:

- measuring the contextual importance of the components;
- measuring the probabilistic importance of the components based on a
5 given corpus;
- measuring the importance of the components based on linguistic knowledge;
- synthesizing the contextual, probabilistic and knowledge based importance measures into a relative importance score for each component; and
10 marking those components having an importance score below a threshold value for removal.

24. The method of generating a summary of claim 23, wherein the contextual importance is measured by:

- identifying a plurality of lexical links of at least one type among the
15 components in a local context in the document; and
- computing a content importance score according to the type and number of lexical links associated with each component.

25. The method of generating a summary of claim 24, wherein the at least one type of lexical links are selected from the group consisting of repetition, inflectional
20 variants, derivational variants, synonyms, hypernyms, antonyms, part-of, entailment, and causative links.

26. The method of generating a summary of claim 23, wherein the probabilistic importance score is determined based on a corpus of human-written summaries.

27. The method of generating a summary of claim 23, wherein the linguistic
25 knowledge operation includes the use of a combined lexicon.

28. The method of generating a summary of claim 22, wherein the sentence combination operation further comprises:

- identifying at least one sentence combination operation;
- establishing at least one rule for applying the sentence combination
- 5 operation; and
- applying the at least one rule to combine at least two extracted sentences.

29. The method of generating a summary of claim 28, wherein the at least one sentence combination operation is selected from the group consisting of add

10 descriptions, aggregations, substitute incoherent phrases, substitute phrases with more general or more specific information, and mixed operations.

30. The method of generating a summary of claim 28, wherein the at least one rule to combine extracted sentences includes replacing a partial name phrase with a full name phrase.

31. The method of generating a summary of claim 28, wherein the at least one rule to combine extracted sentences includes determining if two sentences having a common subject are proximate and whether at least one sentence is marked for reduction then removing the subject of the second sentence and combining with the first sentence using the connective "and."

15

32. A method of identifying correspondence between phrases in a sentence in a summary and phrases in the original document corresponding to the summary comprising:

20

- establishing a plurality of heuristic rules for identifying a cut and paste summarization operation;
- 25 building a probability model based on the heuristic rules; and
- calculating the best solution of the probability model to map a correspondence between the summary phrases and the original phrases.

33. The method of claim 32, wherein the probability model is a Hidden Markov Model.

34. The method of claim 33, wherein a Viterbi algorithm is employed to calculate the best solution.

- 5 35. A corpus for a summarization system comprising:
- a plurality of documents;
 - a plurality of human generated summaries associated with the plurality of documents;
 - 10 a sentence combination subcorpus; and
 - a sentence reduction subcorpus.

36. The corpus of claim 35, wherein the sentence combination subcorpus includes at least one mapping between a summary sentence and at least two original sentences containing phrases in the summary sentence.

- 15 37. The corpus of claim 35, wherein the sentence reduction subcorpus includes at least one sentence pair, each sentence pair having a summary sentence and a corresponding original sentence.

(19) World Intellectual Property Organization
International Bureau



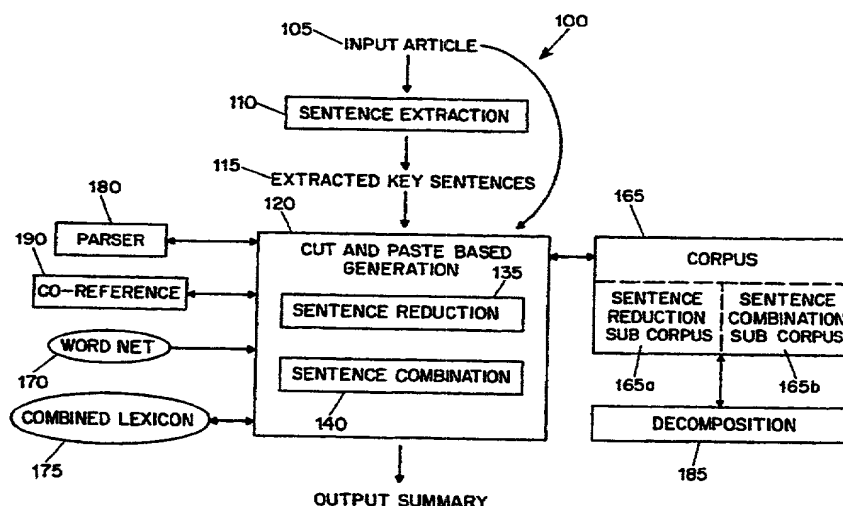
(43) International Publication Date
25 January 2001 (25.01.2001)

PCT

(10) International Publication Number
WO 01/06408 A1

- (51) International Patent Classification⁷: G06F 17/27 (74) Agents: TANG, Henry et al.; Baker & Botts LLP, 30 Rockefeller Plaza, New York, NY 10112-0228 (US).
- (21) International Application Number: PCT/US00/04505
- (22) International Filing Date: 22 February 2000 (22.02.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/120,657 19 February 1999 (19.02.1999) US
- (71) Applicant (for all designated States except US): THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK [US/US]; 116th Street and Broadway, New York, NY 10027 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): MCKEOWN, Kathleen, R. [US/US]; 20 Prospect Road, Wayne, NJ 07470 (US). JING, Hongyan [US/US]; 521 West 112th Street, Apt. 73C, New York, NY 10025 (US).
- (81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:
— With international search report.
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CUT AND PASTE DOCUMENT SUMMARIZATION SYSTEM AND METHOD



(57) Abstract: A summary of an input document is generated by extracting at least one sentence from the document and parsing the extracted sentences into components, such as in a parse tree (110). Sentence reduction processing is performed to mark components which can be removed from the parse trees (135). Sentence reduction can include context importance processing, probabilistic processing, and linguistic knowledge based processing. Probabilistic processing includes identifying sentence combination operations and establishing rules for applying the sentence combination operations to mark the parse trees to merge at least two sentences (140). Sentence combination processing also provides a paste operation to operate on the marked components to effect the indicated removal and combination of sentence components, thereby generating summary sentences for the input document.

2021-01-24 15:50

WO 01/06408 A1

1/8

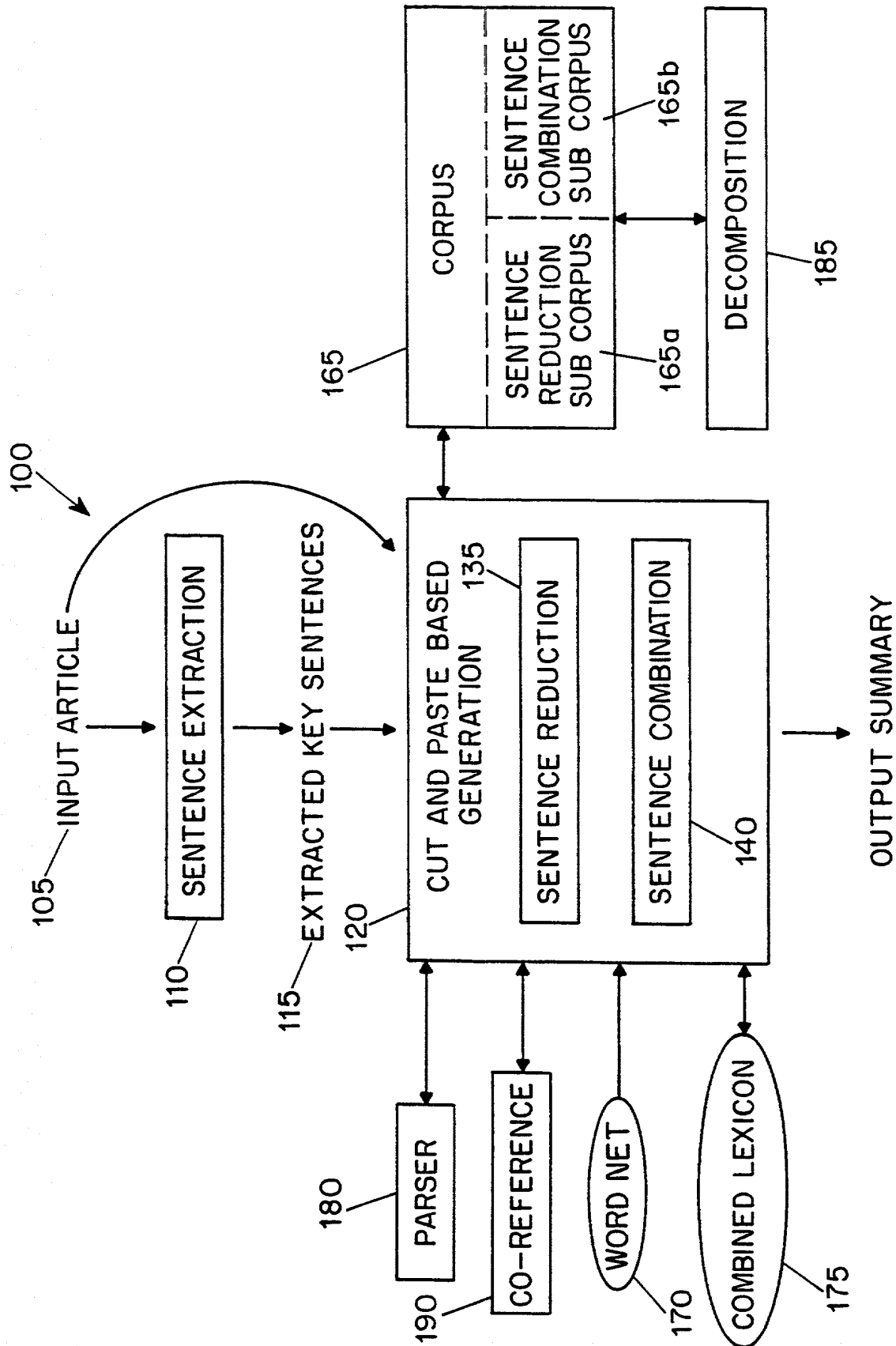
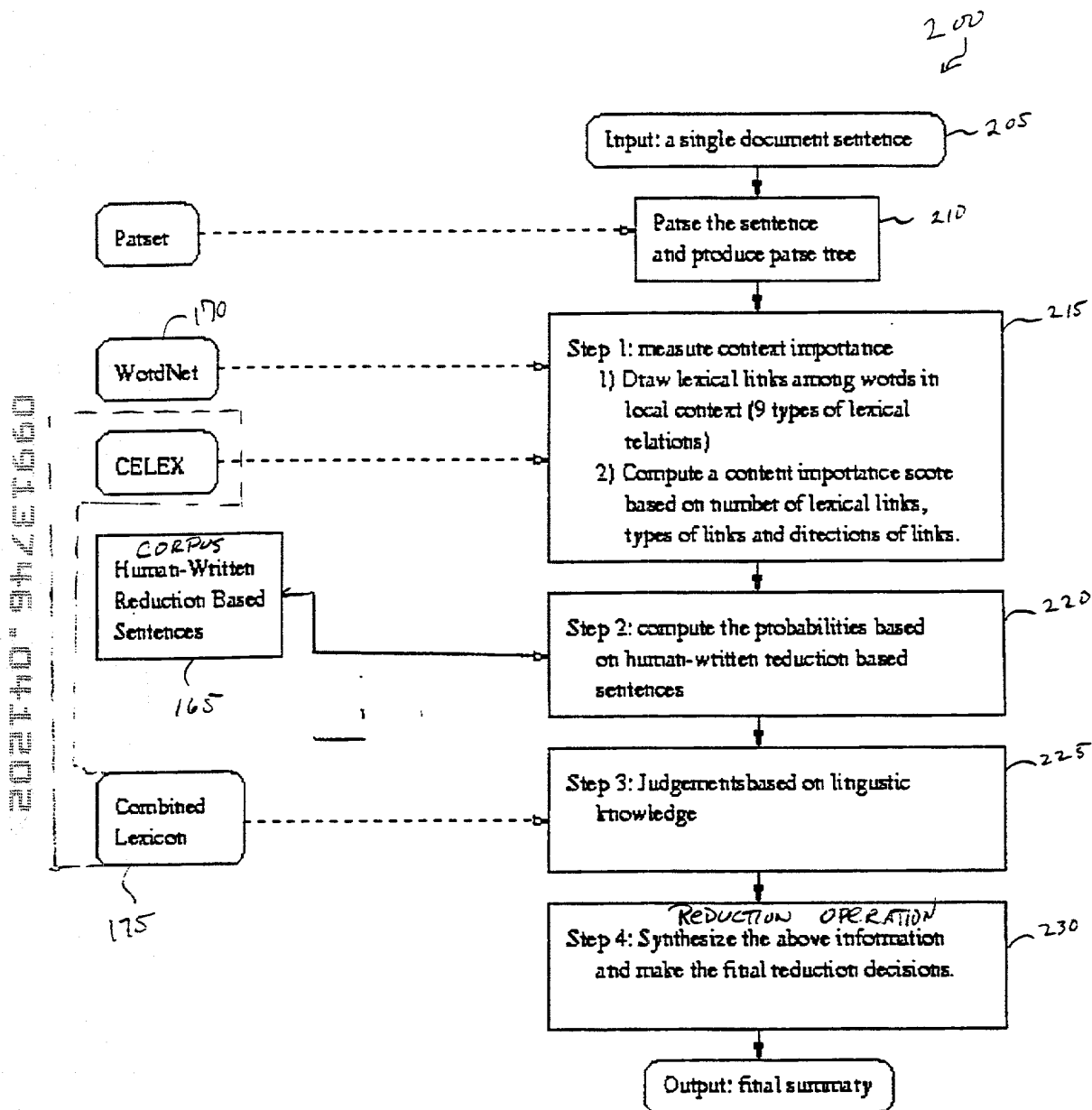


FIG. 1

FIG. 2

2 1 8



3/8

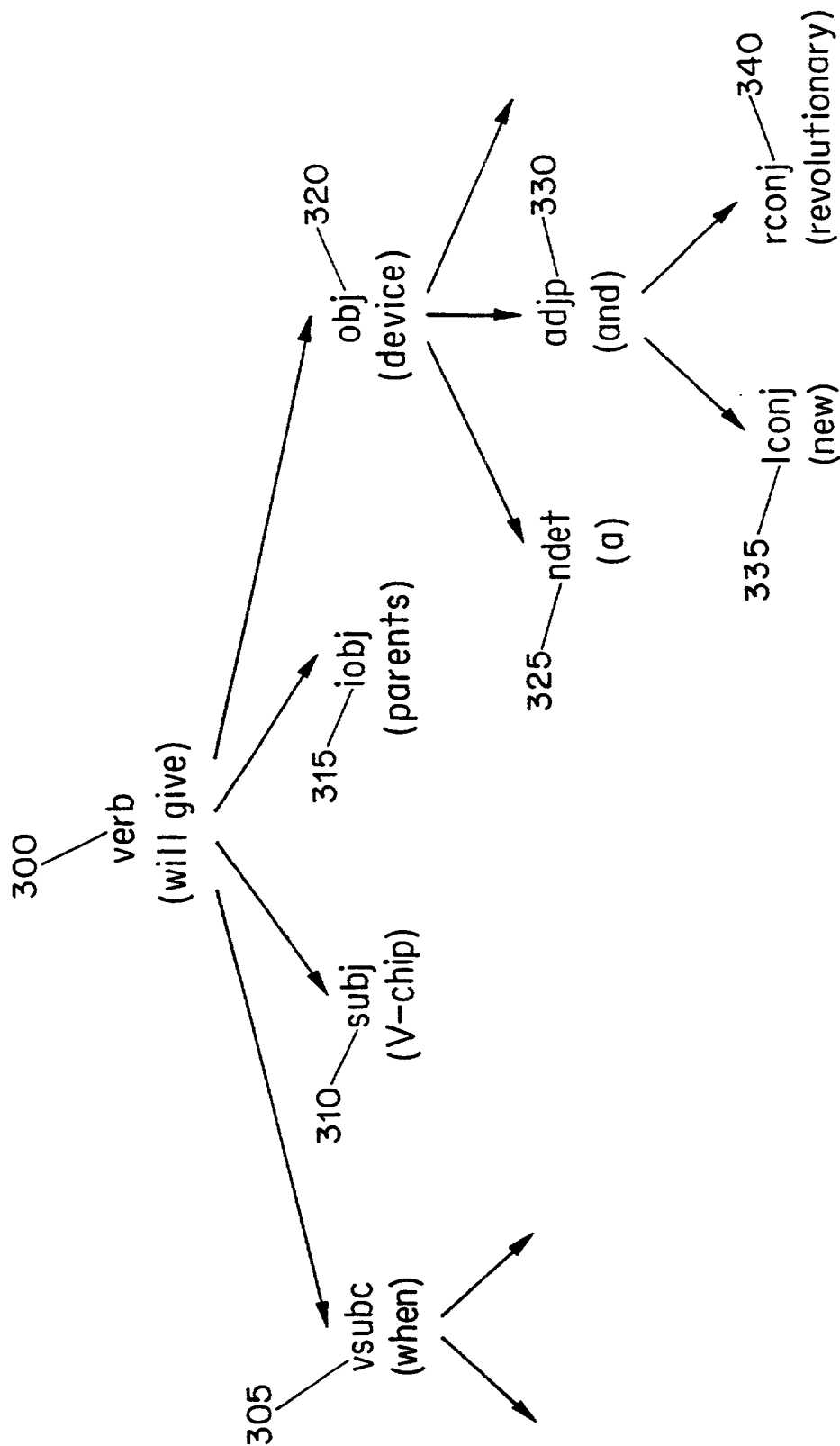


FIG. 3

4/8

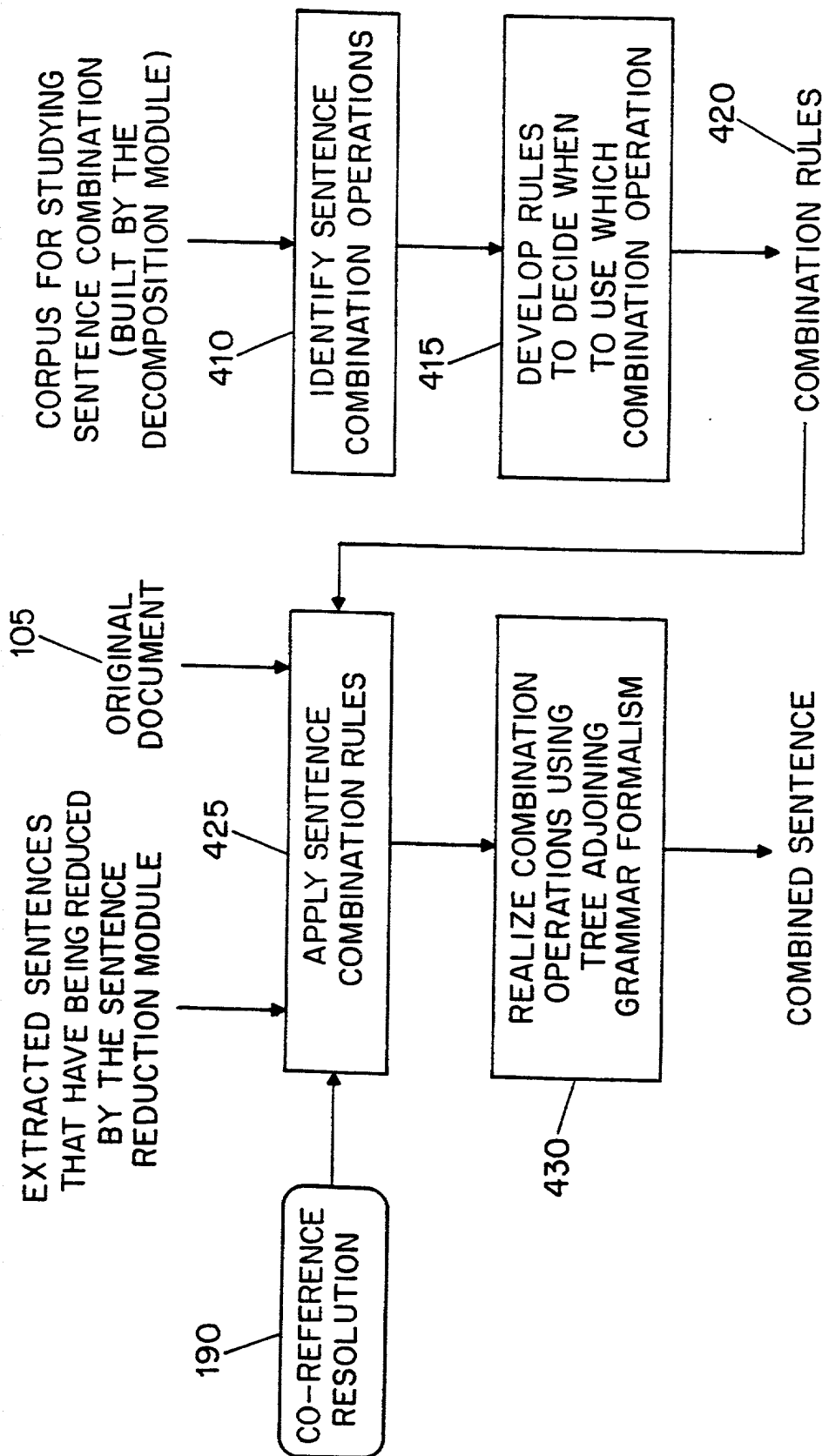


FIG. 4

CATEGORIES	COMBINATION OPERATIONS
510 Add descriptions or names for people or organizations	add description (see Figure 4)
515 Aggregations	add name
	extract common subjects or objects (see Figure 4)
	change one sentence to a clause
	add connectives (e.g., <i>and</i> or <i>while</i>)
	add punctuations (e.g., " ; ")
520 Substitute incoherent phrases	substitute dangling anaphora
	substitute dangling noun phrases
	substitute adverbs (e.g., <i>here</i>)
525	remove connectives
530 Substitute phrases with more general or specific information	substitute with more general information
	substitute with more specific information
	combination of any of above operations (see Figure 2)

FIG. 5

6/8

RULE 1:

IF: ((a person or an organization is mentioned the first time) and (the full name or the full description of the person or the organization exists somewhere in the original article but is missing in the summary))

THEN: replace the phrase with the full name plus the full description

RULE 2:

IF: ((two sentences are close to each other in the original article) and (their subjects refer to the same entity) and (at least one of the sentences is the reduced form resulting from sentence reduction))

THEN: merge the two sentences by removing the subject in the second sentence, and then combining it with the first sentence using connective "and".

FIG. 6

7/8

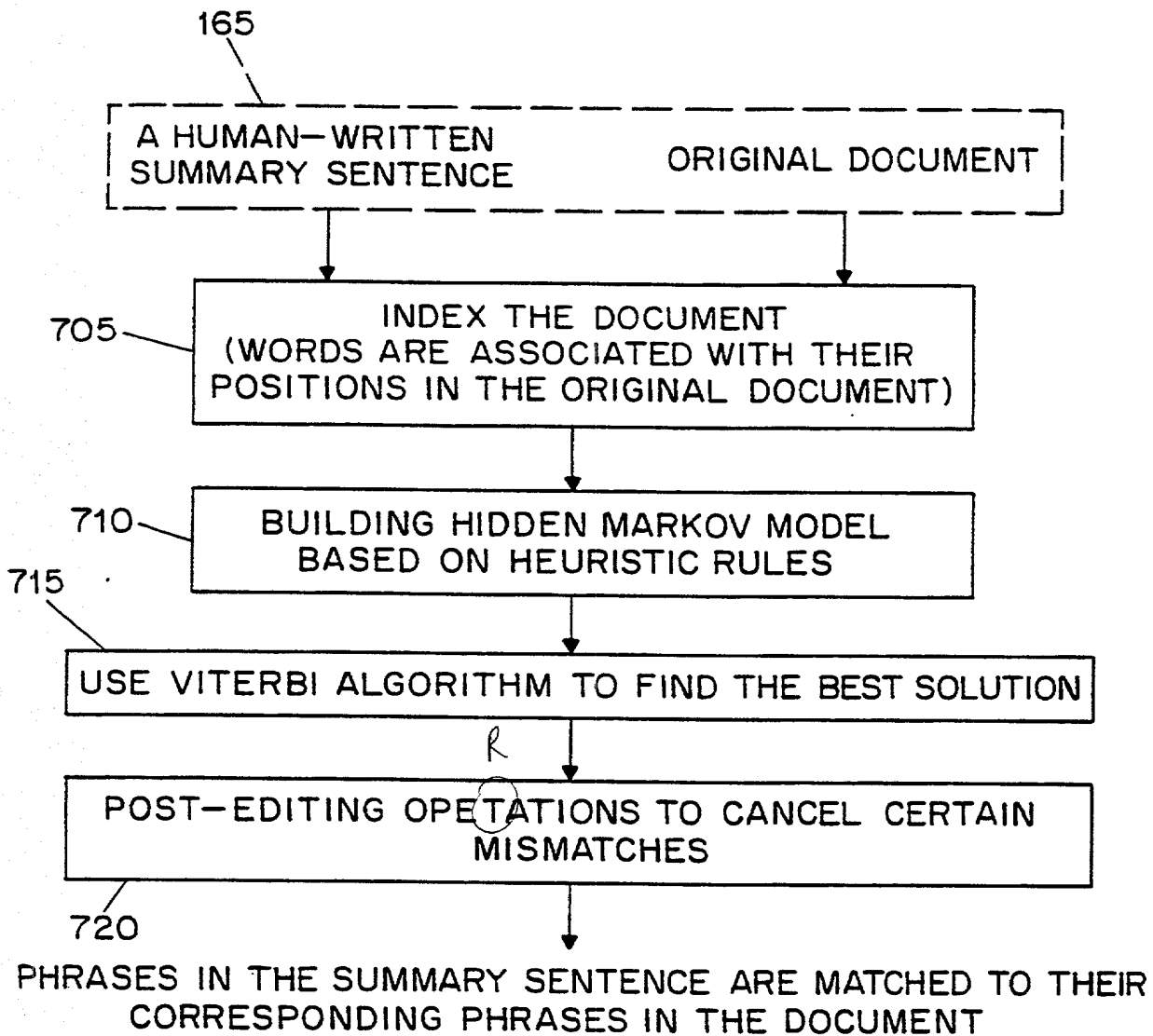


FIG. 7

8/8

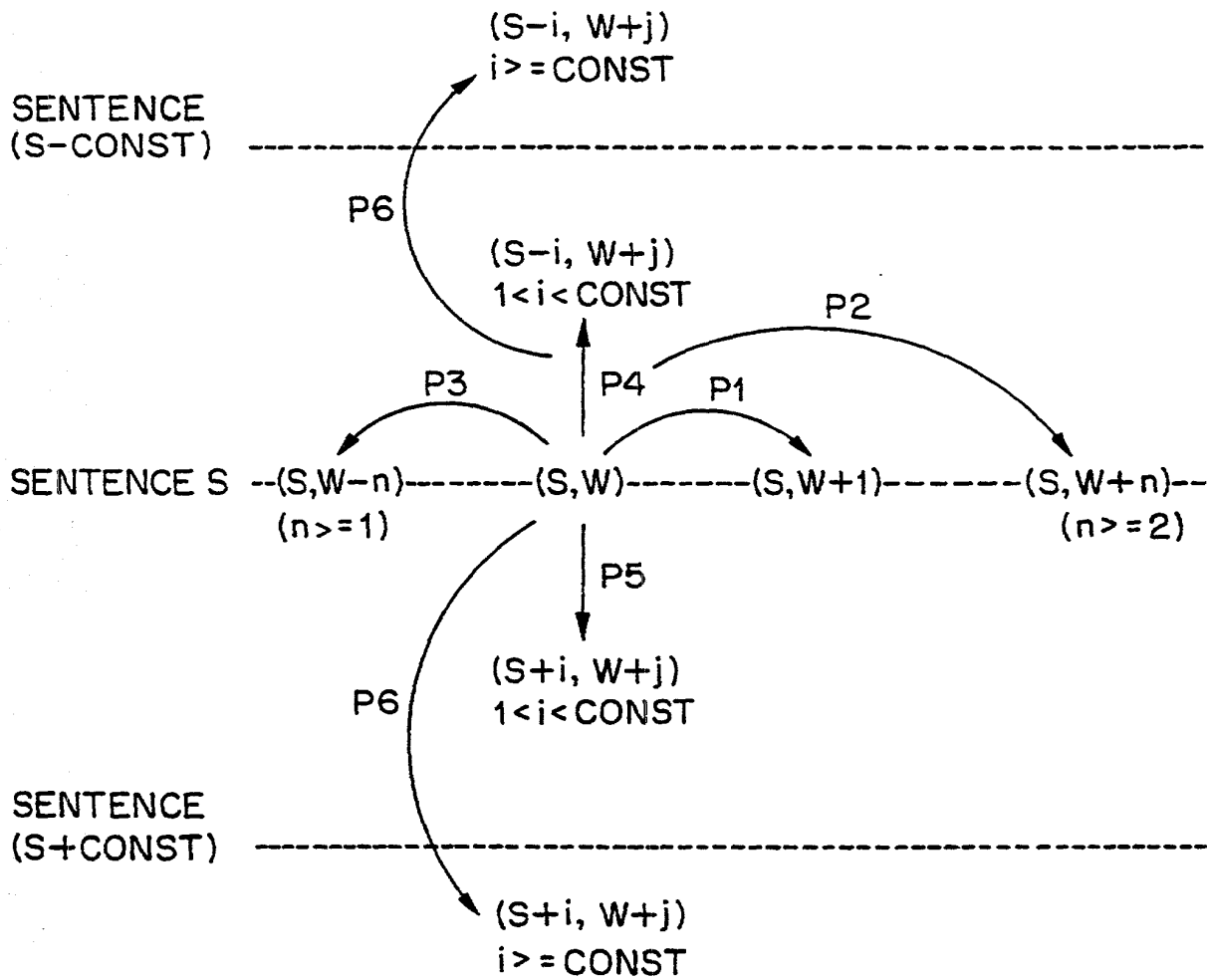
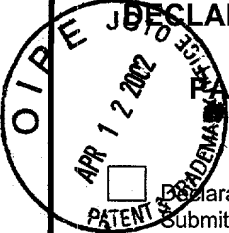


FIG. 8

 <p>DECLARATION FOR UTILITY OR DESIGN PATENT APPLICATION (37 CFR 1.63)</p> <p><input type="checkbox"/> Declaration Submitted with Initial Filing</p> <p>OR</p> <p><input checked="" type="checkbox"/> Declaration Submitted after Initial Filing (surcharge (37 CFR 1.16 (e)) required)</p>	Attorney Docket Number	A32312 PCT-USA
	First Named Inventor	Kathleen R. McKeown
	COMPLETE IF KNOWN	
	Application Number	09/913,746
	Filing Date	August 16, 2001
	Group Art Unit	(Not yet assigned)
	Examiner Name	Stephen Hong

As a below named inventor, I hereby declare that:

My residence, mailing address, and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

CUT AND PASTE DOCUMENT SUMMARIZATION SYSTEM AND METHOD

(Title of the Invention)

the specification of which

☐ is attached hereto

OR

☒ was filed on (MM/DD/YYYY) 08/16/2001 as United States Application Number or PCT International

Application Number 09/913,746 and was amended on (MM/DD/YYYY) (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment specifically referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56, including for continuation-in-part applications, material information which became available between the filing date of the prior application and the national or PCT international filing date of the continuation-in-part application.

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or (f), or 365(b) of any foreign application(s) for patent, inventor's or plant breeder's rights certificate(s), or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent, inventor's or plant breeder's rights certificate(s), or any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Number(s)	Country	Foreign Filing Date (MM/DD/YYYY)	Priority Not Claimed	Certified Copy Attached?	
				YES	NO
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

☐ Additional foreign application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto:

DECLARATION — Utility or Design Patent Application

Claim for Benefit of Prior U.S. Provisional Application(s)

I hereby claim the benefit under Title 35, United States Code, § 119(e) of any United States provisional application(s) listed below:

Provisional Application Number	Filing Date

Claim for Benefit of Earlier U.S./PCT Application(s) under 35 U.S.C. 120

(complete this part only if this is a divisional, continuation or C-I-P application)

I hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s) or PCT international application(s) designating the United States of America that is/are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior application(s) in the manner provided by the first paragraph of Title 35, United States Code § 112, I acknowledge the duty to disclose information as defined in Title 37, Code of Federal Regulations, Section 1.56 which occurred between the filing date of the prior applications(s) and the national or PCT international filing date of this application:

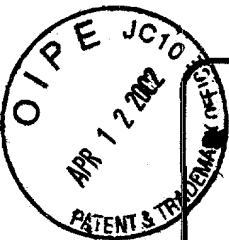
Application Number	Filing Date	Status (patented, pending, abandoned)
PCT/US00/04505	February 22, 2000	Pending

BAKER BOTTS LLP

Attorney Docket Number A32312 PCT-USA

DECLARATION — Utility or Design Patent Application

Direct all correspondence to: <input checked="" type="checkbox"/>		Customer Number or Bar Code Label	21003		OR	<input checked="" type="checkbox"/> Correspondence address below
Name						
Address						
City			State		ZIP	
Country			Telephone		Fax	
<p>I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.</p>						
NAME OF SOLE OR FIRST INVENTOR :			<input type="checkbox"/> A petition has been filed for this unsigned inventor			
Given Name (first and middle [if any])			McKeown Family Name or Surname			
Inventor's Signature			Kathleen R. McKeown		Date	
Wayne Residence: City			NJ State		U.S.A. Country	
20 Prospect Road			NJ State		U.S.A. Country	
Mailing Address			07470 ZIP		U.S.A. Country	
NAME OF SECOND INVENTOR:			<input type="checkbox"/> A petition has been filed for this unsigned inventor			
Given Name (first and middle [if any])			Jing Family Name or Surname			
Inventor's Signature			Hongyan Jing		Date 3/27/2002	
New York Residence: City			NJ State		U.S.A. Country	
621 West 112th Street, Apt. 73C			NJ State		Peoples Republic of China Citizenship	
Mailing Address			28 GALES DR. APT 2		U.S.A. Country	
New York City			NJ State		40025-07974 ZIP	
New Providence			NJ State		U.S.A. Country	
<input type="checkbox"/> Additional inventors are being named on the _____ supplemental Additional Inventor(s) sheet(s) PTO/SB/Q2A attached hereto.						

BAKER BOTTS LLPPlease type a plus sign (+) inside this box → **POWER OF ATTORNEY OR
AUTHORIZATION OF AGENT**

Application Number	09/913,746
Filing Date	August 16, 2001
First Named Inventor	Kathleen R. McKeown
Group Art Unit	(Not yet assigned)
Examiner Name	Stephen Hong
Attorney Docket Number	A32312 PCT-USA

I hereby appoint:

☒ Practitioners at Customer Number

21003

Place Customer
Number Bar Code
Label here

☐ Practitioner(s) named below:

Name	Registration Number

as my/our attorney(s) or agent(s) to prosecute the application identified above, and to transact all business in the United States Patent and Trademark Office connected therewith.

Please change the correspondence address for the above-identified application to:

☒ The above-mentioned Customer Number.

OR

☐ Firm or
Individual Name

Address

Address

City

State

Zip

Country

Telephone

Fax

I am the:

☒ Applicant/Inventor.

☐ Assignee of record of the entire interest. See 37 CFR 3.71.
Statement under 37 CFR 3.73(b) is enclosed. (Form PTO/SB/96).
SIGNATURE of Applicant or Assignee of Record

Name

Hongyan Jing

Signature

Hongyan Jing

Date

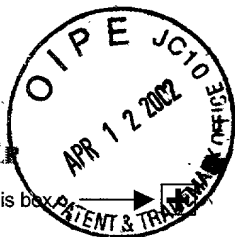
3/27/2002

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required. Submit multiple forms if more than one signature is required, see below*.

☐ *Total of _____ forms are submitted.

2002140" 3745650

BAKER BOTTS LLP



Please type a plus sign (+) inside this box

**POWER OF ATTORNEY OR
AUTHORIZATION OF AGENT**

Application Number	09/913,746
Filing Date	August 16, 2001
First Named Inventor	Kathleen R. McKeown
Group Art Unit	(Not yet assigned)
Examiner Name	Stephen Hong
Attorney Docket Number	A32312 PCT-USA

I hereby appoint:

☒ Practitioners at Customer Number

21003

Place Customer
Number Bar Code
Label here

☐ Practitioner(s) named below:

Name	Registration Number

as my/our attorney(s) or agent(s) to prosecute the application identified above, and to transact all business in the United States Patent and Trademark Office connected therewith.

Please change the correspondence address for the above-identified application to:

☒ The above-mentioned Customer Number.

OR

☐ Firm or
Individual Name

Address

Address

City

State

Zip

Country

Telephone

Fax

I am the:

☒ Applicant/Inventor.

☐ Assignee of record of the entire interest. See 37 CFR 3.71.
Statement under 37 CFR 3.73(b) is enclosed. (Form PTO/SB/96).

SIGNATURE of Applicant or Assignee of Record

Name

Kathleen R. McKeown

Signature

Kathleen R. McKeown

Date

4/3/02

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required. Submit multiple forms if more than one signature is required, see below*.

☐ *Total of _____ forms are submitted.